

SEQUENCE LISTING

## (1) GENERAL INFORMATION:

## (i) APPLICANT:

(A) NAME: Fred Hutchinson Cancer Research Center, Inc.  
 (B) STREET: 1100 Fairview Avenue North, Mailstop c2M-027  
 (C) CITY: Seattle  
 (D) STATE: Washington  
 (E) COUNTRY: USA  
 (F) POSTAL CODE (ZIP): 98109

(A) NAME: Thomas Spies  
 (B) STREET: 2429 E. Aloha  
 (C) CITY: Seattle  
 (D) STATE: Washington  
 (E) COUNTRY: USA  
 (F) POSTAL CODE (ZIP): 98112

(A) NAME: Veronika Spies  
 (B) STREET: 2429 E. Aloha  
 (C) CITY: Seattle  
 (D) STATE: Washington  
 (E) COUNTRY: USA  
 (F) POSTAL CODE (ZIP): 98112

(ii) TITLE OF INVENTION: CELL STRESS REGULATED HUMAN MHC CLASS I GENE

(iii) NUMBER OF SEQUENCES: 16

## (iv) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk  
 (B) COMPUTER: IBM PC compatible  
 (C) OPERATING SYSTEM: PC-DOS/MS-DOS  
 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30 (EPO)

## (vi) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/029,044  
 (B) FILING DATE: 29-OCT-1996

## (2) INFORMATION FOR SEQ ID NO: 1:

## (i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 11722 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

CACTGCTTGA GCCGCTGAGA GGGTGGCGAC GTCGGGGCCA TGGGGCTGGG CCCGGTCTTC	60
CTGCTTCTGG CTGGCATCTT CCCTTTTGCA CCTCCGGGAG CTGCTGCTGG TGAGTGGCGT	120
TCCTGGCGGT CCTCGGCGGA GCGGGAGCAG TGGGACGTTT CCGGGGGTCG GGTGGGTAGC	180
GGCGAGCGCT GTGCGGTCAG GGC GGCGGCTC CTGTGCCCTG TCGGTGGCGC AGGGAGCTGG	240

ACGCGGCCCG TTACCGCCAC ACTTCAGCCC TGCTTCCCCG TCACTTTTCA GTCCTCCTCG	300
GGATCGCGCA TCACCTGCAC TTTCTGGTCT CCTCCTGCTC TTTCTCTCCT CGCGTCTCCT	360
CCGCTTCCTC TCACTTTTCG GACAAACCAG TCCTTCTGAG GCCCATGGGT TCCCGGGCTG	420
CCTCCGGGGC TGCTCCTGTG AATGGCATTC GAGTGGCCTT CCAGCGCGGC CACTGAAGCA	480
GCCACAACCC CCGTGCTCG GGGCGGCTCT CAGGTCCCTG AAGTCCTGTC CTCTCCCGGA	540
GCCGACGTGT TCTCAGCTCC TGGGCCGAG CTCCTGGAGT AGGGGCCCTC CTTTCTCGGG	600
ACCCGGAGCT GGTGCTTCCT GCTGCTGTGG GGACTGTGGG GGGTCCTGAC TCTCAAGCTG	660
AGGGGTGGA GTCTGCAGGC TCCGGGCAGA GGATTCTTCC TGCGACTTCT CTCATCCCCA	720
GCTCATTCTC CCCTCGCCTC TGGCTCCGAG GGTCTCTCC TCTCTCTCAT CCCACCCCTA	780
CTAATGACCA GTGATCTAAG GACACCAGAT TCCCTCTCAC CTCCTCCCTG CCCATCTCAG	840
GGCCCGCTGA GTCCTTTTGC CCTCCCAGCT CCCTGCTACC CCTCCTGTG TGCTGTTCTC	900
TGATCCATTT CTAGGGTGTC CTCTGCCCTC ATCCCCTGTC CCCGCCACCG AAGTCCCTCC	960
TGCACCCCTT ATGGGCCTTT CCTACAAGCA GCCTTCACCC AGTGCTGCCC CTATGCCTCC	1020
CCGTTCCCAA ATGTCCCTGA CTCTAACTTT CTGGTGCTGC CTTTATCCG GGGGGGTCTT	1080
CCCTCCATCC CACTCCCCTC CAGACCCCCA AGGGGAACCC TGATGCTAAT GGCAGTTGGG	1140
CCTTAGGCAG GGCAGAGGC AGCGCAGATG CCCCCTCCCC TCCAGTGAG ATGCCTGTTT	1200
TGGACCCTGC CTCATTGTGG CCCCTTCCCC ACTCCTTCAT CCTCAGCCTC ACCCTCTTGA	1260
GGACCCACC CTCCAGCCCA CAGGTGCTGG ACCATCCCTC CCTGGTCCCT CCGCCCCCTT	1320
CCACCTTGGG ACCTTGTGCT GCTCCTATCT CTTGCCCAGC TGCCTTGGGC CCTCAGCACG	1380
TTCTCATCTT TCAGTGGGAA AGTGGGAGTG CTGGAGCATA TGACAGTGCT GAGCATCTTT	1440
CCCAAGCCCC ACCCTCCCC AGAGCACCT CCCCTCCTGT CCTCACCTA CCCCAGTTC	1500
TCCCACAGTC ACTCCTGCCC CATGCTCATG CCGCCCTCCA GTTCTTGCTC TGCCCATCTC	1560
CCCTCCCCAA CCCAGACCTA AAACAGGCTG TTGGGCCAAC TGTTCTTGA CTTCTCTCT	1620
TTTCTTTTGG TTCCTTGACC CCAGTGGGCT CTCACTCCCC ACACCGCATA TCTAAAATCT	1680
GTTTTGCCTG CTCTTGGGGT GCCACTGCTC CCCCTCCAGC ATTACTCCTT TTGGCAGGTC	1740
CTTCCTCAGG CTGAGAATCT CCCCCTCTAC CTTGGTTTTT TCTCTCTGGC CAGCACCCCC	1800
ACTCCTTGCT TTGTTTTTAA TTTTAACTT TTGTTTGGGT ACGTAGTAGA TATATATGTA	1860
TATATTTATG GGGTACATGG GATATTTTGA CACAGGCCA CAATATGTAA TAATCACATC	1920
AGGGTAAATG GGTATATCA CAACAAGCAT TTATCCTTTC TTTGTGCTAC AAACAATCCC	1980

ATTATGCTCT TTCAGTTATT TTTAAATGTA CAATAAATTA TTGTTGACTG TACTCACCCCT	2040
GCTGTGCTAT CTAGTAGATC TTATTCATTC TAATTATATT TTTGTACCCA TTATTAACCA	2100
TCCCTGCTCC CCCACTCCCC ACTACCCTTC TCAGCCTCTG GTAATCATCA TTCTATTGTC	2160
TCTCCCCATG AGGTCCATTG TTTTAAATTT TGGCTGCCAC AAATAAGTGA GAACATGCAA	2220
AGTTTGTCTG TCTGGGCCTG GGGCTTATTT CACTTCACAG GATGACCTCC AGTTCTTTGC	2280
AAATGACACG ATGGCTGAAT AGTTCTCCAC ATACACATGT ACACCACATT TTCTTTATCC	2340
ATGCGTCTGT TGATGGACAC TTAGATTGCT TGCAGATCTT GGCTACTTTG AATAGTGCTG	2400
CAATAAACAT GGAAAAGTAG ATAGCTCTTT AATATACCGA TTCCTTTCT TTGGAGTATA	2460
TGCCTAACAG TGGGAGTGCT GGAGCATATG ACAGCTCTAT TGTATTTTGA GTTTTGGAA	2520
GAACCTCCAC ATTGTTTCCC ATAGTGGTTG TACTAGTTTA CGTCCCACC AACAGTGAC	2580
ATCCTCACCA GCATTCCCTA TTTCTACATC CTCGCCAGCA TTCCTTATTG CCTGTCTTCT	2640
GGATAAAGC CAGTTTATCT GGGGTGGGAT GTTATCTCGT AGGAGTTTTG ATTTGCCTTC	2700
ATCTGTTGAC GAATGATGTT GAGCACCTTT TCATATACCT GTTTGCCATT TATATGTCTT	2760
CTTTTGAGAA ATGACTATTC AGATCTTTTC TCATTTTAA ATTGGATTAT TATATTTTTT	2820
TTCTATAGT TGTTGAGCT CCTTATATGT TTCAGTTACT GATCCTTTGT CAGATGAATA	2880
GTTTGAAAAT ATTTTCTCCC ATTCTTGGAT GGTCTCTTCA TTTTGTTTAT TGTTTCCTTT	2940
GCTGTGCAGA AGCCTTTTGA CTTGATATGA TCCCATTAT GCAATTTTAC TTTGGTTACC	3000
TGTGCTTG TG GGGTATTACT TTAAAAATCT TTGCCAGTC CAATATCCTA GAGAGTTTCC	3060
CCAATGTTTT CTGTATAGT TTCATAGTTT GAGGTCATAG ATTTACATCT TTAATCCACT	3120
TTGATTGAT TTTTGTATAT GGTGAAAGAC AGGGTCTAGT TTCATTCTTC TGCATAAGGA	3180
TATCTAGTTT CCCCAGCACC ATTTTGAAG AGACTCTCCT TTGCCAATGT GTGTTCTTGG	3240
TACCTTTGTT GGAAATGAGT TTAGTGTAGA TGTATGGAAT TGTTTCTGGG TTCTCTATTC	3300
TGTTTCATTG GTCTGTGTGT CTGTTTTTAT GCCAGTATCA TGCTGTTTTG GTTACTGTAG	3360
CTCTGTAGTA TAATTTGAAG TCAGATAATG TGATTCCTCT AGTTTTGTTC ATTTTGCTCA	3420
GGATAGCTTT ATCTATTCTG GTTTTTTTGT GGTTCATAT GCATTTTAGG ATTATTTTTA	3480
TTATTTCTGT GAAGAATGTC ATTAGTGTTT TGATAGGGAT TGCATTGAAT CTGTAGATTA	3540
CTTTGGGTAG TATGGATATT TCAACAAAAC TGATTCTTCC AATCCATGAA CGTGGACTAT	3600
CTTTTCATT TTTTGTGTCC TTCAATTTTT TGCATCAGTG TTTTGTGTTT TTGGTTTTTG	3660
AGATGGAGTT TCACTCTTGT TGCCAGGCT AGAATGCAAG GGTGTGATCT TGGCTCACCG	3720

CAACCTCCGC CTCCCAGGTT CAAGCTATTC TTCTGCCTCA GCCTCCCAAG TAGCTGGGAT	3780
TACAGGCATG TGCCACTGTG CCTGGCTAAT TTTCTATTTT TATTAGAGAT GGGGTTTCTC	3840
TATGTTGGCC AGGCTAGTCT TGAACCTCTG ACCTCAGGTG ATCCACCTGC CTCGGCCTCC	3900
CAAAGTGCTG GGATTACAGG CATGAGCCAC CACGCCCAGC CACATCACTG TTTTATAGTT	3960
TTTATTGGAG AGGTCTTTCA CTTCTTCAGT TAGGTTTATT CCTCAGTATT TTATTTTATT	4020
TGTAGCTATT GTAAATGGGA TTCGTTTCTT GATTTCTTTT TCAGATTATT TGCTGTTAGC	4080
ACTGATTTTT GCATGTTGAT TTTGTATCCT GCAACTTTAC TGAATTTGTT CTTCAGTTCT	4140
AATGGTTTTT TGGTGGAGTC TTTAGGTTTT TCCAAATATC AGACCACATG ATCTGCAAAC	4200
AAGGATAATT TGACTTCTTC TTTTCCAGTT TTAATGCCCT TTCTTTCTTT CTCCTGTCTG	4260
ATTGCTCTAG TTAGGATCTG CAGTACTGTG TTGCATAACT GTGGTAAAAT TAGTCATCCT	4320
TGTCTTATTC CAGATCTTAG AGAAAAGGCT TTCAGTTTTT CCCCATTGAG TATGTTACTA	4380
GCTGTGAGTT TGTCATATAT GGCTTTTATT ATATTGAGGT CTGTTCTTGG TATACTTAGT	4440
TTTTTGAGAG TTTTATCAT GAAGGGATGT TGAATTTATC AAATGCTTTT TCAGTATCAA	4500
TTGAATGATA CTGGCTTTTG TCCTTTATTC TGTGATATG ACGTATTACA TTGATTGATT	4560
TGTGTATGTT AAATCATCCT TGCATACCTG GAATACATTC CACTTGCTCA TAAAGAATGA	4620
TCTTTTTTAA TGTATTGTTG AATGTGGTTT GCTAGTATTT CCTTGACGAT TTTTGCATCG	4680
GTGTTTCATCA GGGATATAGG CCTGTAGTTT TCTTTTTTAT GATGTGTCTT TGCCTGGTTT	4740
TTGTATCAGG ATATTCCTGG CTTGTAAAA TGAGTTTGA AGTATTCCCT CCTCCTCTAT	4800
TTTTCAGAAC AGTTTGAATA GGA CTGACAT ATGTTGTTCT TAAAAAGTTT AATTGTGGTA	4860
AATTATACAT TACATAAATT TTA CTGTTTT AACCCTTTT AAGTGTATAC TCGGTGGCAT	4920
TAGATACATT CACATTTTTG TGCAACCCAA AACTCTGTGC CCATTAATCG GTAAC TCCC	4980
ATTCTCCCT ACCTCTGGCC CCTGGTAACC ACCATTCTAC TTTTGTTC TATGAATTTG	5040
ACCACTCTAG GTACCTCATT TAAGCAGAAT CATGTAATGT TTGTCTTTTT GTTCTGGCT	5100
TATTTCACTT ATAATATTTT TGAGGTTTCG TGGGCACAGT GGCTCACGCC TGGATTCCA	5160
GCACTTTGGG AGGCTGAAGC AGGTGGATCA CCTGAGTTTC GGAGTTGAA ACCAGCCTGG	5220
CCAACATGGT GAAACCCCAT CTCTACTAAA AATAATAAAA GTTAGCCGGG CGTGATGGCG	5280
GGTGCCTGTA ATCCCAACTA CTTGGGAGGC TGAGGCAGGA GAATCGCTTG AATCCGGGAA	5340
GTGGAGGTTG CAGTGAGCTG AGATCAGGCC ACTGCACTCC AGCCTGGGCA ACAAGAGTGA	5400
AATTCCATCT CCAAAAAAAA AAAATAAAC AATAATAATA ATAATATTTT TGAGGTTTAT	5460

CCAAGTTGTA GTATGGGTCA GAATTTTCATT CCTTTTAAGG ATGGATAATA CTCATTATAT	5520
GTATGTACCA CATCTTGGTT ATCCATCCCT CAGACAATGG ACACTTGGGT TACTTCTACC	5580
TTTTGGATAT TGGCAAATAT TTCATTTCTT TTGGGTATAT ATTTATTTCC TTTGGGTATT	5640
TCTTTTGGGT ATATATCCAG AAATAGAAGC AGTACACAGG GGCTTCATTT TCTCTGTCTC	5700
TTTGCCAACC TTGCTCTGTG TGTGTGTGTA TGTGTGTGTG TAGGTGTGTG ATAACAGCCA	5760
TCCTGATTGG TTTCAGGTGG CATCTCATTG TGGTTTGGAT TTGCATTTTC CTAATGAGTG	5820
CTGATATTGA GCATCTTTTC ATGTGTTTGT TGATCATTTG TAATTTTCTT TGAAGAATTG	5880
GCCATTTAAG TCTTTTGCCC ATTTTTTCCC CCACATAGCT TCTCTTATCA GATATATGAC	5940
TTGCAATATT TATTTTCATT CGGGGTTGAT TGCTTTTCA CTCTGATTGT GCCCTTTGAT	6000
GCATAGATGT TTTGAATTTT CATCAGTCTA CTTTGTGAGT TCTTTCTATT CTATCTGTGC	6060
TTTGGTGTCA TATCCATGAA AGCACTGTCA AATCCTATGT CATGAACATT ATCCCCAATG	6120
TTTGCTTCTA AGAAATTTT AGGTTTTAGT TCTTGAGTGT AGAGTTTAGG TCTTTGATTG	6180
ATTTTGAGTT AATTTTGTGA TATAGTGCAA ATTAAGGGTC CAATTTTATT TTAACACCCC	6240
CTGCCCCCAG AACTATTTGC TGAAAAGATC AACTGACTCT TTGTCACCTG CTCACCCCAG	6300
TGGACACTAG CTGTTCCATC CAATTGCTGT CCTGGGGCCT TGTCATGCTA CTCTTCCACT	6360
TTGAACCCAA GCCCACACCG TTCGTTGCTC CCCTCTGGGA TACTGACCCC ACTATAAACT	6420
TCTCTGGGGC TACAACCTTC CTACCCTTTG TGCCTCATGA CCACCCCCTC CCTTGTCCCC	6480
GCCATGCCCA TGATGAGTCT CTTCTCGAGG CAGCTCCCCT TGCCTCCATC TCACCCTCAG	6540
CCTATGCACC ACAGCCACAC TGGACATGGG TCCCTCTGAG CCTGAGTCCC TTCCCATTCC	6600
CACCATCTCC TCTGGCAAGA CCTTCCTTCC ACCACCTTCA TGCTCCTCCC TTGCCCCTGC	6660
AGGGCAGCCT CTCCCCTTGG CCCCTATTCC CTTAGGGGGC TTGTGGCCAC CCAGTCCTTG	6720
CACCTGGCCT ACAAGTTTGC CATCTTCATT CCCCCTTCTT CTGTTTCATCA GCCCCCTCCT	6780
CTATCCTCCC ACCCTCACAG TTTTCTTTGT ATATGAAATC CTCGTTCTTG TCCCTTTGCC	6840
CGTGTGCATT TCCTGCCCCA GGAAGGTTGG GACAGCAGAC CTGTGTGTTA AACATCAATG	6900
TGAAGTTACT TCCAGGAAGA AGTTTCACCT GTGATTTCTT CTTCCCCAGA GCCCCACAGT	6960
CTTCGTTATA ACCTCACGGT GCTGTCCTGG GATGGATCTG TGCAGTCAGG GTTTCTTGCT	7020
GAGGTACATC TGGATGGTCA GCCCTTCCTG CGCTATGACA GGCAGAAATG CAGGGCAAAG	7080
CCCCAGGGAC AGTGGGCAGA AGATGTCCTG GGAAATAAGA CATGGGACAG AGAGACCAGG	7140
GACTTGACAG GGAACGGAAA GGACCTCAGG ATGACCCTGG CTCATATCAA GGACCAGAAA	7200

GAAGGTGAGA	GTCGGCAGGG	GCAAGAGTGA	CTGGAGAGGC	CTTTTCCAGA	AAAGTTAGGG	7260
GCAGAGAGCA	GGGACCTGTC	TCTTCCCACT	GGATCTGGCT	CAGGCTGGGG	GTGAGGAATG	7320
GGGGTCAGTG	GAACCTCAGCA	GGGAGGTGAG	CCGGCACTCA	GCCCACACAG	GGAGGCATGG	7380
GGGAGGGCCA	GGGAGGCGTA	CCCCCTGGGC	TGAGTTCCTC	ACTTGGGTGG	AAAGGTGATG	7440
GGTTCGGGAA	TGGAGAAGTC	ACTGCTGGGT	GGGGGCAGGC	TTGCATTCCC	TCCAGGAGAT	7500
TAGGGTCTGT	GAGATCCATG	AAGACAACAG	CACCAGGAGC	TCCCAGCATT	TCTACTACGA	7560
TGGGGAGCTC	TTCCTCTCCC	AAAACGTGGA	GACTGAGGAA	TGGACAGTGC	CCCAGTCCTC	7620
CAGAGCTCAG	ACCTTGGCCA	TGAACGTCAG	GAATTTCTTG	AAGGAAGATG	CCATGAAGAC	7680
CAAGACACAC	TATCACGCTA	TGCATGCAGA	CTGCCTGCAG	GAACCTACGGC	GATATCTAGA	7740
ATCCAGCGTA	GTCCTGAGGA	GAAGAGGTAC	GGACGCTGGC	CAGGGGCTCT	CCTCTCCCTC	7800
CAATTCTGCT	AGAGTTGCCT	CACCTCCAAG	ATGTGTCCAG	GGAAACCCTC	CCTGTGCTAT	7860
GGATGAAGGC	ATTTCCTGTT	GGCACATCGT	GTCCTGATTT	TCCTCTATTG	TTAGAGCCAC	7920
TGGATAAAGA	CAGTGGGTCA	GGGACTGGAC	CATCCAGTGT	TGTAATCAGG	GCAAGTAGAG	7980
GACCTTCCGA	CAGAATCCTG	AGCCTGTGGT	GGGTGTCAGG	CAGGAGAGGA	AGCCTTCAGG	8040
GCCAGGGCTG	CCCCCTCTGC	CTCCCAGCCT	GCCCATCCTG	GAGAGTTCCC	TCCTGGCCCC	8100
ACAACCCAGG	AGTCCACCCC	TGACATCCCC	CTCCTCAGCA	TCAATGTGGG	GATCCAGAG	8160
CCTGAGGCCA	CAGTCCCAAG	GCCCATCCTC	CTGCCAGCCT	GGAAGAACTG	GGCCCCAGAG	8220
TGAGGACAGA	CTTGCAGGTC	AGGGGTCCCG	GAGGGCTTCA	GCCAGAGTGA	GAACAGTGAA	8280
GAGAAACAGC	CCTGTTCTCT	TCCCCTCCTT	AGAGGGGAGC	AGGGCTTCAC	TGGCTCTGCC	8340
CTTTCTTCTC	CAGTGCCCCC	CATGGTGAAT	GTCACCCGCA	GCGAGGCCTC	AGAGGGCAAC	8400
ATCACCGTGA	CATGCAGGGC	TTCCAGCTTC	TATCCCCGGA	ATATCACACT	GACCTGGCGT	8460
CAGGATGGGG	TATCTTTGAG	CCACGACACC	CAGCAGTGGG	GGGATGTCCT	GCCTGATGGG	8520
AATGGAACCT	ACCAGACCTG	GGTGGCCACC	AGGATTTGCC	AAGGAGAGGA	GCAGAGGTTC	8580
ACCTGCTACA	TGGAACACAG	CGGGAATCAC	AGCACTCACC	CTGTGCCCTC	TGGTGAGCCT	8640
AGGGTGACCC	TGGAGAGGGT	CAGGCCAGGG	TAGGGACAGC	AGGGATGGCT	GTGGCTCTCT	8700
GCCCAGTGTA	TAACAAGTCC	CTTTTTTTC	GGGAAAGTGC	TGGTGCTTCA	GAGTCATTGG	8760
CAGACATTCC	ATGTTTCTGC	TGTTGCTGCT	GCTGCTGCTG	CTATTTTGT	TATTATTATT	8820
TTCTATGTCC	GTTGTTGTAA	GAAGAAAACA	TCAGCTGCAG	AGGGTCCAGG	TGAGAAAAGC	8880
GGGCAGTTTC	TGGAGATGGT	AAGGCCCTG	TCTGGGCAGT	AGGGTCCCCT	CATTGCTCCT	8940

GCAAAGATAG GCATGTTGGT GACAAGGCTT CTGTAACAGG GGATGAAAGT TGGGGAATTT	9000
GGGAAGGGAA TGGGGGCAGC ATCTCCATCT ACACCCATAA GTGCTGCCCA AGCGAGGGTC	9060
AAACGCCCAG CTGTGGCATC TTCCTGCTGC AGGTGAGGAG TGGGCAGCAG GGAGGGCTGC	9120
GGCGCCTGCT CTGTCCCCAT CCCGGTCTCT GTGTCTCTTG GACTCACTAG GGCGCATCCA	9180
GGTGGGGTGA GCTGGGAATC ACGTGCTGAA TGCTGAGGGC CTGGATGATC ACGGCCTCAG	9240
AGGGAGCAAA TAGTAAAGGC AGCTGTGATC TGGGGAGGGC CAGAAACTGG AGAGGAATCT	9300
GAGGAGAGGC GGTGCCCTTA TTCCCTTCCT CTCTGCATCC CCCTCCCCTG TTTCTCCAGC	9360
CATCGGGGCG GACACCGAGA AAAAGACCTA TGAGGCCAG CCTGGGGGCC CTGCCTGTGT	9420
AGCCCTTTGG AGACCCCTAG TAACAGGGAG GGTCTGAGC ACACATGGCC ATCTCTGTCC	9480
ACTGTGCAGC TCCCCATGCA CCTCCTCCAG GAGCTTTCTT GGGGTTGTCG TGTCTCTGC	9540
ACCATTGAG GCCCTACTCT TTCCAGGTC CCACGGCTG GCCTCCCTGA GTTCTTGCA	9600
GATGACATGG ATGAGTAGAT AAGCAGATGT CCCTGGGCCA TTTGAGGAGT GGGGCCAGC	9660
CCCTCATCAG GGCAGCTGTG GTCCCTGTTT TCATCCTACC TCCGAGTGT TTCTTCTCCA	9720
GTCCCTGAGG GACACAGTCC TCAGGGCCCA TGTTTTTGGG GATTTAATCT GTGCTCTGTG	9780
GCCTCACCTT GCCTTCCCTG AGCCAATTTC CCTTTCTAAA GGTGGTCACT GCCTGGTAAG	9840
TTTGAGTAGTAA GGGACGGTCA GAATCATTTT CCCTACAGTC AGGTGTTTG ATGGGGGATG	9900
AAAAGAGACA GCAGGAAGTT TTGTGTTTCT GCAAAGACAG AAGCAGTTCA GGCGACAGTA	9960
AGAGGCTGGG GTGTCCAGGA GGGTGTGTCT GGCAGTAGGG TCGCTGGTTT CTCATCCTTG	10020
AACCTAATTG CACTGTCAGT CGGCCCTCA GGCCTGAGCA GATGGGAAGG TTTGTCCCCT	10080
GCCCTGCAGC AAGAGGGCCC TGTCCAGGAG GCACCCACAA CAGAGGCAGT GCAGGTCTGT	10140
GGTCACTCCT ACTCTCACCT GTGGCGTCTC CCGTAGAGGG ATTGTCAGTT CTGGTTCCCT	10200
GTGGGCAGGA ATGGTTTCCT CATAGGTCAC TGGAGTTTGG GCCAGGAAAA GAGTATGAAG	10260
TTCATGTGGC AGTTTCTCAA AATTCTGCT TTCAATGTTG ATGTCCAGTA AAGATATTCG	10320
TAATTTGAGC TCTATAATCT TAATAGGATT TCCTCTAATA TTGTGAAGCA TATTATATGA	10380
AACAGGAACA CAAATTTCTC AAAATTCCTG CGATGTCCAA TAAAGATTTT CATAATTTCA	10440
GCTCTGCAAT CTTAATAGGA TTTCTAATA CTGTAAAGCA TATTAAATGA AACAGGAAGT	10500
CAAATTTGGA GCCCCCTCTC CAGGAGGTTT TGTGTGGAGA TGGTGGCTGT GGCAGTGGCA	10560
GTTCCAGGT GCAGAGGGTG GGCAGAGGCA GCCTCAGGCT AAGGGGTCTC CCCTACTCCA	10620
CATGGAGAAA ATCCCTTGTA GGTTGCAAGG GCAGTGCCCG GGTGGAATCC CTGCTAGGGA	10680

CAGAGCAGGA AGGCCTCGCA GCCTCACCAA GCAGCAGCCC TGGGGTGGAG CTGCGTTTCC 10740  
 AGGGTTAAGC GGACCAGGCA GGAGTAGCGG TTAACAAGA GCAGGTCACA GGCTTGGGTT 10800  
 GTGAGGGTCA GGAGAGGCCA GGCCTCCTCG AGCAAGGTGG GGGTCCCAGG GTCAGGTCAG 10860  
 GTGCAGATCC TGTGGCAGCC ACGTCTTTCC ATGCTGGGCC TGCTGGGCC CCCAGGCTTC 10920  
 CTGATGGGGT CCCAGTTAG GAGCTGCCTG CTCAGGGCTG GGAGGGGAGG AGCACTGAGC 10980  
 TGCAGATAGA GGGCAGAGCC CACAGTGGGC AGGGCCTGCC CTGGTGTGTA GGTGCCTCTG 11040  
 CAGGAGAGGA GGGCCTGGGG ACTGAGAGCA AGGGTCAGGG CCTCTCTTTG GGGAGGCCTC 11100  
 TCACTGTAAC AGGACTGGTC AGGCCTGAGA GGAGGGCACT GGGTTCCCTC TTGGGTCTTG 11160  
 TCCTTTAGTC TTGGGGCCCT TTCCCTCCCT GCACGATGAG TGGTGGGCAC AGGGCACGGG 11220  
 CTGATGTTGA TGGAGTGATG GGAGGGAAC TGCAGGGGCT GGGAAAAGCA AGGAGGGAGG 11280  
 AAGAAAAAAG TGGGGGCCTC ATCTTCCCTC AGAGAAAGGG CAAATCTGGT TTTGGAGCAA 11340  
 CTGAAGAGAG AAAAGTCCCC AGGGAATAAA CACAACACTG CACCCAGTGG AGCATTACC 11400  
 CATTTCCCTC TTTTCTCCAG AGCTCGTGAG CCTGCAGGTC CTGGATCAAC ACCCAGTTGG 11460  
 GACGAGTGAC CACAGGGATG CCACACAGCT CGGATTTCAG CCTCTGATGT CAGCTCTTGG 11520  
 GTCCACTGGC TCCACTGAGG GCGCCTAGAC TCTACAGCCA GCGGGCTGGA ATTGAATTCC 11580  
 CTGCCTGGAT CTCACAAGCA CTTTCCCTCT TGGTGCCTCA GTTTCCTGAC CTATGAAACA 11640  
 GAGAAAATAA AAGCACTTAT TTATTGTTGT TGGAGGCTGC AAAATGTTAG TAGATATGAG 11700  
 GCATTTCAG CTGTGCCATA TT 11722

## (2) INFORMATION FOR SEQ ID NO: 2:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 385 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

Met Gly Leu Gly Pro Val Phe Leu Leu Leu Ala Gly Ile Phe Pro Phe  
 1 5 10 15  
 Ala Pro Pro Gly Ala Ala Ala Glu Pro His Ser Leu Arg Tyr Asn Leu  
 20 25 30  
 Thr Val Leu Ser Trp Asp Gly Ser Val Gln Ser Gly Phe Leu Ala Glu  
 35 40 45



Val His Leu Asp Gly Gln Pro Phe Leu Arg Tyr Asp Arg Gln Lys Cys  
 50 55 60  
 Arg Ala Lys Pro Gln Gly Gln Trp Ala Glu Asp Val Leu Gly Asn Lys  
 65 70 75 80  
 Thr Trp Asp Arg Glu Thr Arg Asp Leu Thr Gly Asn Gly Lys Asp Leu  
 85 90 95  
 Arg Met Thr Leu Ala His Ile Lys Asp Gln Lys Glu Gly Leu His Ser  
 100 105 110  
 Leu Gln Glu Ile Arg Val Cys Glu Ile His Glu Asp Asn Ser Thr Arg  
 115 120 125  
 Ser Ser Gln His Phe Tyr Tyr Asp Gly Glu Leu Phe Leu Ser Gln Asn  
 130 135 140  
 Val Glu Thr Glu Glu Trp Thr Val Pro Gln Ser Ser Arg Ala Gln Thr  
 145 150 155 160  
 Leu Ala Met Asn Val Arg Asn Phe Leu Lys Glu Asp Ala Met Lys Thr  
 165 170 175  
 Lys Thr His Tyr His Ala Met His Ala Asp Cys Leu Gln Glu Leu Arg  
 180 185 190  
 Arg Tyr Leu Glu Ser Ser Val Val Leu Arg Arg Arg Val Pro Pro Met  
 195 200 205  
 Val Asn Val Thr Arg Ser Glu Ala Ser Glu Gly Asn Ile Thr Val Thr  
 210 215 220  
 Cys Arg Ala Ser Ser Phe Tyr Pro Arg Asn Ile Thr Leu Thr Trp Arg  
 225 230 235 240  
 Gln Asp Gly Val Ser Leu Ser His Asp Thr Gln Gln Trp Gly Asp Val  
 245 250 255  
 Leu Pro Asp Gly Asn Gly Thr Tyr Gln Thr Trp Val Ala Thr Arg Ile  
 260 265 270  
 Cys Gln Gly Glu Glu Gln Arg Phe Thr Cys Tyr Met Glu His Ser Gly  
 275 280 285  
 Asn His Ser Thr His Pro Val Pro Ser Gly Lys Val Leu Val Leu Gln  
 290 295 300  
 Ser His Trp Gln Thr Phe His Val Ser Ala Val Ala Ala Ala Ala  
 305 310 315 320  
 Ala Ile Phe Val Ile Ile Ile Phe Tyr Val Arg Cys Cys Lys Lys Lys  
 325 330 335  
 Thr Ser Ala Ala Glu Gly Pro Glu Leu Val Ser Leu Gln Val Leu Asp  
 340 345 350

Gln His Pro Val Gly Thr Ser Asp His Arg Asp Ala Thr Gln Leu Gly  
 355 360 365

Phe Gln Pro Leu Met Ser Ala Leu Gly Ser Thr Gly Ser Thr Glu Gly  
 370 375 380

Ala  
 385

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2380 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

GGGCCATGGG GCTGGGCCGG GTCCTGCTGT TTCTGGCCGT CGCCTTCCCT TTTGCACCCC	60
CGGCAGCCGC CGCTGAGCCC CACAGTCTTC GTTACAACCT CATGGTGCTG TCCCAGGATG	120
AATCTGTGCA GTCAGGGTTT CTCGCTGAGG GACATCTGGA TGGTCAGCCC TTCCTGCGCT	180
ATGACAGGCA GAAACGCAGG GCAAAGCCCC AGGGACAGTG GGCAGAAGAT GTCCTGGGAG	240
CTAAGACCTG GGACACAGAG ACCGAGGACT TGACAGAGAA TGGGCAAGAC CTCAGGAGGA	300
CCCTGACTCA TATCAAGGAC CAGAAAGGAG GCTTGCAATC CCTCCAGGAG ATTAGGGTCT	360
GTGAGATCCA TGAAGACAGC AGCACCAGGG GCTCCCGGCA TTTCTACTAC GATGGGGAGC	420
TCTTCCTCTC CCAAACCTG GAGACTCAAG AATCGACAGT GCCCCAGTCC TCCAGAGCTC	480
AGACCTTGGC TATGAACGTC ACAAATTTCT GGAAGGAAGA TGCCATGAAG ACCAAGACAC	540
ACTATCGCGC TATGCAGGCA GACTGCCTGC AGAAACTACA GCGATATCTG AAATCCGGGG	600
TGGCCATCAG GAGAACAGTG CCCCCATGG TGAATGTCAC CTGCAGCGAG GTCTCAGAGG	660
GCAACATCAC CGTGACATGC AGGGCTTCCA GCTTCTATCC CCGGAATATC ACACTGACCT	720
GGCGTCAGGA TGGGGTATCT TTGAGCCACA ACACCCAGCA GTGGGGGGAT GTCCTGCCTG	780
ATGGGAATGG AACCTACCAG ACCTGGGTGG CCACCAGGAT TCGCCAAGGA GAGGAGCAGA	840
GGTTCACCTG CTACATGGAA CACAGCGGGA ATCACGGCAC TCACCCTGTG CCCTCTGGGA	900
AGGTGCTGGT GCTTCAGAGT CAACGGACAG ACTTTCATA TGTTCCTGCT GCTATGCCAT	960
GTTTTGTTAT TATTATTATT CTCTGTGTCC CTTGTGCAA GAAGAAAACA TCAGCGGCAG	1020
AGGGTCCAGA GCTTGTGAGC CTGCAGGTCC TGGATCAACA CCCAGTTGGG ACAGGAGACC	1080
ACAGGGATGC AGCACAGCTG GGATTTCAGC CTCTGATGTC AGCTACTGGG TCCACTGGTT	1140

CCACTGAGGG CGCCTAGACT CTACAGCCAG GCGGCCAGGA TTCAACTCCC TGCCTGGATC 1200  
 TCACCAGCAC TTTCCCTCTG TTTCTGACC TATGAAACAG AAAATAACAT CACTTATTTA 1260  
 TTGTTGTTGG ATGCTGCAAA GTGTTAGTAG GTATGAGGTG TTTGCTGCTC TGCCACGTAG 1320  
 AGAGCCAGCA AAGGGATCAT GACCAACTCA ACATTCCATT GGAGGCTATA TGATCAAACA 1380  
 GCAAATTGTT TATCATGAAT GCAGGATGTG GGCAAACTCA CGACTGCTCC TGCCAACAGA 1440  
 AGGTTTGCTG AGGGCATTCA CTCCATGGTG CTCATTGGAG TTATCTACTG GGTCACTAG 1500  
 AGCCTATTGT TTGAGGAATG CAGTCTTACA AGCCTACTCT GGACCCAGCA GCTGACTCCT 1560  
 TCTTCCACCC CTCTTCTGTC TATCTCCTAT ACCAATAAAT ACGAAGGGCT GTGGAAGATC 1620  
 AGAGCCCTTG TTCACGAGAA GCAAGAAGCC CCCTGACCCC TTGTTCCAAA TATACTCTTT 1680  
 TGTCTTTCTC TTTATTCCCA CGTTCGCCCT TTGTTCAATC CAATACAGGG TTGTGGGGCC 1740  
 CTTAACAGTG CCATATTAAT TGGTATCATT ATTTCTGTTG TTTTGTGTTT TGTTTTGTTT 1800  
 TTTGTTTTTG AGACAGAGTC TCACTCGTCA CCCAGGCTGC AGTTCACTGG TGTGATCTCA 1860  
 GCTCACTGCA ACCTCTGCCT CCCAGGTTCA AGCACTTCTC GTACCTCAGA CTCCCGATAG 1920  
 CTGGGATTAC AGACAGGCAC CACCACACCC AGCTAATTTT TGTATTTTTT GTAGAGACGG 1980  
 GGTTCGCCA AGTTGACCAG CCCAGTTTCA AACTCCTGAC CTCAGGTGAT CTGCCTGCCT 2040  
 TGGCATCCCA AAGTGCTGGG ATTACAAGAA TGAGCCACCG TGCCTGGCCT ATTTTATTAT 2100  
 ATTGTAATAT ATTTTATTAT ATTAGCCACC ATGCCTGTCC TATTTTCTTA TGTTTTAATA 2160  
 TATTTTAATA TATTACATGT GCAGTAATTA GATTATCATG GGTGAACTTT ATGAGTGAGT 2220  
 ATCTTGGTGA TGAATCCTCC TGACCAGCCC AGGACCAGCT TTCTTGTCAC CTTGAGGTCC 2280  
 CCTCGCCCCG TCACACCGTT ATCGATTACT CTGTGTCTAC TATTATGTGT GCATAATTTA 2340  
 TACCGTAAAT GTTTACTCTT TAAATAAAAA AAAAAAAAAA 2380

## (2) INFORMATION FOR SEQ ID NO: 4:

## (i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 383 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

## (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Met Gly Leu Gly Arg Val Leu Leu Phe Leu Ala Val Ala Phe Pro Phe  
 1 5 10 15  
 Ala Pro Pro Ala Ala Ala Ala Glu Pro His Ser Leu Arg Tyr Asn Leu  
 20 25 30

Met Val Leu Ser Gln Asp Glu Ser Val Gln Ser Gly Phe Leu Ala Glu  
 35 40 45  
 Gly His Leu Asp Gly Gln Pro Phe Leu Arg Tyr Asp Arg Gln Lys Arg  
 50 55 60  
 Arg Ala Lys Pro Gln Gly Gln Trp Ala Glu Asp Val Leu Gly Ala Lys  
 65 70 75 80  
 Thr Trp Asp Thr Glu Thr Glu Asp Leu Thr Glu Asn Gly Gln Asp Leu  
 85 90 95  
 Arg Arg Thr Leu Thr His Ile Lys Asp Gln Lys Gly Gly Leu His Ser  
 100 105 110  
 Leu Gln Glu Ile Arg Val Cys Glu Ile His Glu Asp Ser Ser Thr Arg  
 115 120 125  
 Gly Ser Arg His Phe Tyr Tyr Asp Gly Glu Leu Phe Leu Ser Gln Asn  
 130 135 140  
 Leu Glu Thr Gln Glu Ser Thr Val Pro Gln Ser Ser Arg Ala Gln Thr  
 145 150 155 160  
 Leu Ala Met Asn Val Thr Asn Phe Trp Lys Glu Asp Ala Met Lys Thr  
 165 170 175  
 Lys Thr His Tyr Arg Ala Met Gln Ala Asp Cys Leu Gln Lys Leu Gln  
 180 185 190  
 Arg Tyr Leu Lys Ser Gly Val Ala Ile Arg Arg Thr Val Pro Pro Met  
 195 200 205  
 Val Asn Val Thr Cys Ser Glu Val Ser Glu Gly Asn Ile Thr Val Thr  
 210 215 220  
 Cys Arg Ala Ser Ser Phe Tyr Pro Arg Asn Ile Thr Leu Thr Trp Arg  
 225 230 235 240  
 Gln Asp Gly Val Ser Leu Ser His Asn Thr Gln Gln Trp Gly Asp Val  
 245 250 255  
 Leu Pro Asp Gly Asn Gly Thr Tyr Gln Thr Trp Val Ala Thr Arg Ile  
 260 265 270  
 Arg Gln Gly Glu Glu Gln Arg Phe Thr Cys Tyr Met Glu His Ser Gly  
 275 280 285  
 Asn His Gly Thr His Pro Val Pro Ser Gly Lys Val Leu Val Leu Gln  
 290 295 300  
 Ser Gln Arg Thr Asp Phe Pro Tyr Val Ser Ala Ala Met Pro Cys Phe  
 305 310 315 320  
 Val Ile Ile Ile Ile Leu Cys Val Pro Cys Cys Lys Lys Lys Thr Ser  
 325 330 335

Ala Ala Glu Gly Pro Glu Leu Val Ser Leu Gln Val Leu Asp Gln His  
                   340                                  345                                  350

Pro Val Gly Thr Gly Asp His Arg Asp Ala Ala Gln Leu Gly Phe Gln  
                   355                                  360                                  365

Pro Leu Met Ser Ala Thr Gly Ser Thr Gly Ser Thr Glu Gly Ala  
                   370                                  375                                  380

(2) INFORMATION FOR SEQ ID NO: 5:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

ACTGGGGAAC AAGGTTTATA TGAGA

25

(2) INFORMATION FOR SEQ ID NO: 6:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 24 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

TGTCACCCGT CTTCTACAGG ACCC

24

(2) INFORMATION FOR SEQ ID NO: 7:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

GGGGCCATGG GGCTGGG

17

(2) INFORMATION FOR SEQ ID NO: 8:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 17 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

ATCTGAGATG TCGGTCC

17

(2) INFORMATION FOR SEQ ID NO: 9:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CGTTCTTGTC CCTTGCCCG TGTGC

25

(2) INFORMATION FOR SEQ ID NO: 10:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 25 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

AACCCTTCCC TTACCCCGT CGTAG

25

(2) INFORMATION FOR SEQ ID NO: 11:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 45 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

TATGTAAAC GACGGCCAGT TTCACCTGTG ATTTCCTCTT CCCCA

45

(2) INFORMATION FOR SEQ ID NO: 12:

- (i) SEQUENCE CHARACTERISTICS:
  - (A) LENGTH: 45 base pairs
  - (B) TYPE: nucleic acid
  - (C) STRANDEDNESS: single
  - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:

GGTCTTTTCA ATCCCGTCT CTCGTCCAGT ATCGACAAAG GACAT

45

(2) INFORMATION FOR SEQ ID NO: 13:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 40 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TATGTAAAAC GACGGCCAGT TTCGGAATG GAGAAGTCAC

40

(2) INFORMATION FOR SEQ ID NO: 14:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 40 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

CGAGAGGAGA GGGAGGTAA CCAGTATCGA CAAAGGACAT

40

(2) INFORMATION FOR SEQ ID NO: 15:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 40 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TATGTAAAAC GACGGCCAGT GTTCCTCTCC CTCCTTAGA

40

(2) INFORMATION FOR SEQ ID NO: 16:

- (i) SEQUENCE CHARACTERISTICS:  
 (A) LENGTH: 41 base pairs  
 (B) TYPE: nucleic acid  
 (C) STRANDEDNESS: single  
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:

AAAAAGTCCC TTTCACGACC ACCAGTATCG ACAAAGGACA T

41